Arif-Uz-Zaman, Kazi, Cholette, Michael E., Li, Fengfeng, Ma, Lin, & Karim, Azharul

(2015)
A data fusion approach of multiple maintenance data sources for real-world reliability modelling. In
10th World Congress on Engineering Asset Management, 28-30 September 2015, Tampere Hall, Tampere, Finland.

A data fusion approach of multiple maintenance data sources for real-world reliability modelling

Kazi Arif-Uz-Zaman[1], Michael E. Cholette, Fengfeng Li, Lin Ma, Azharul Karim

Abstract A central tenet in the theory of reliability modelling is the quantification of the probability of asset failure. In general, reliability depends on asset age and the maintenance policy applied. Usually, failure and maintenance times are the primary inputs to reliability models. However, for many organisations, different aspects of these data are often recorded in different databases (e.g. work order notifications, event logs, condition monitoring data, and process control data). These recorded data cannot be interpreted individually, since they typically do not have all the information necessary to ascertain failure and preventive maintenance times.

This paper presents a methodology for the extraction of failure and preventive maintenance times using commonly-available, real-world data sources. A text-mining approach is employed to extract keywords indicative of the source of the maintenance event. Using these keywords, a Naïve Bayes classifier is then applied to attribute each machine stoppage to one of two classes: failure or preventive. The accuracy of the algorithm is assessed and the classified failure time data are then presented. The applicability of the methodology is demonstrated on a maintenance data set from an Australian electricity company.

1 Introduction

Companies typically keep data about maintenance of assets in event/maintenance notifications. These data have significant potential to provide asset managers with a rich set of information about the operation of their assets, including their reliability. However, asset data are typically-collected in a "one-size-fits-all" approach focusing on maintenance record keeping rather than reliability modelling and analysis (Louit et al., 2009). In many organizations, data are

[1] K. Arif-Uz-Zaman (✉) · M. E. Cholette · F. Li · L. Ma · A. Karim
Science & Engineering Faculty, Queensland University of Technology (QUT), 2 George St, Brisbane, QLD-4001, Australia.
e-mail: arif.kaziarifuzzaman@hdr.qut.edu.au

recorded describing maintenance actions conducted on the asset. These data typically are: 1) Work Orders /Notifications (WONs), and 2) Downtime Data (DD).

A WON is a record of every action associated with maintenance (including inspection, repair, replace etc.) without specifying if the maintenance is reactive or preventive. A WON tells us if the work is a "defect" or "urgent" but it does not tell us if this constitutes a "failure", i.e. if it stops the operation of the asset. On the other hand, DD contains asset stoppage information without stating whether the downtime is planned or unplanned. Thus, each dataset is incomplete from a reliability modelling point of view, where we need to know *both* when the asset is down and if this downtime was unplanned. Moreover, because the notification entries are made by humans and entered in lay language with little standardization, the variation in the input is extremely large.

Thus, a significant research question arises as: how typically-available asset data can be utilized for reliability models? Few efforts have been made regarding this issue. For example, Bastos et al. (2014) and Jeon and Sohn (2015) develop statistical data extraction methods to extract failure-related information from their chosen datasets. Alkali et al. (2009) used hourly readings of motor current to determine whether the mills were running or not and assumed all downtime was related to failure. Most of the methods usually used failure times which were already available to databases. However, in many cases, required information is buried in various data sets in both numerical and text formats. This complication renders traditional data mining tools unusable.

In this paper we develop a novel method to the extract information required for reliability model using the free text available in data sources. The method presented can link between data available (WON and DD) and information required (failure times) for reliability modelling. The method analyses WONs to construct a keyword dictionary using text descriptions which is in turn used to classify each DD event as a failure or preventive (preventive) maintenance event.

2 Information Extraction Methodology

The overall approach is summarized here which can be seen in Fig. 1. The basic idea is to use the WON free text to construct a classifier using words that the organization typically uses to describe urgent and unexpected maintenance. However, the WONs do not contain reliable downtime information. Thus the keyword dictionary and classifier are applied to the free text of the DD to associate each event with a failure or preventive maintenance action.

2.1 Data Selection and WON Labelling

WONs usually contain information regarding all types of work, planned or unplanned. In order to train a text classifier, WONs need to be labelled as failure and non-failure types. We define failure as unplanned maintenance work requires immediate downtime.

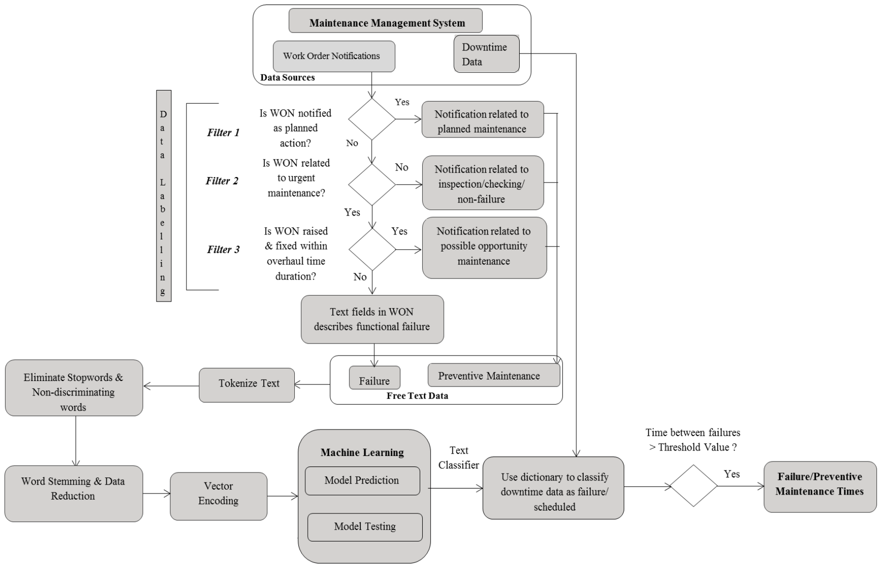

Figure 1 Methodology to extract failure and preventive maintenance information

According to this definition, unplanned maintenance events in WONs are candidate failure data (Filter 1, Fig. 1). However, not all unplanned downtimes are failures. Some unplanned WONs are issued to periodically monitor anomalies or schedule and prioritize preventive maintenance actions during the next planned stoppage. Thus, we select only WONs that are urgent for further analysis (Filter 2 in Fig. 1). In addition, any WONs that are both raised and fixed while the asset has been down are classified as preventive maintenance (Filter 3 in Fig. 1). The overall filter process is shown in Fig. 1 (data labelling). Hence, WONs that have high urgency/priority likely contain language that personnel use to describe failure.

2.2 Data Cleaning & Construction of Keyword Dictionary

The free text from WON (that labelled with two of the classes: failure and preventive maintenance) will then be used to construct a keyword dictionary. After data selection and labelling the free text in the WONs are used to construct a keyword dictionary. Usually, maintenance data contain a large proportion of valuable and interesting information in text formats. For example, description of maintenance work, failure modes, types of maintenance and many more. Since, these free texts are the source of useful information; these can be used to classify the data. But before that, text cleaning is necessary to remove unwanted space, numbers, punctuation and, most importantly, non-discriminating words. At the beginning of cleaning process, all the free text are transformed into lower case followed by removing numbers, punctuation and extra spaces in between the words.

A common practice when analysing text data is to remove filler words such as “to”, “and”, “where”, “or”, “when”, etc. These are known as *stop words*. Apart from that, some keywords are considered to be common but not useful in discriminating between the classes (failure and preventive maintenance here), which need to be eliminated. Text cleaning transforms the raw text into a representation known as *bag-of-words*. This ignores the orders that terms appear in rather than simply provides a variable indicating whether the term appears at all. It is then necessary to transform the terms and sentences into a form that machine learning algorithms can understand. This can be done by splitting the cleaned text documents into individual words, which is called *tokenization*. A token is the single element of text string (keyword). The classifier requires data in the form of table where each row contains a document and each column presents a keyword (Here keywords are the all words within the dictionary) (Noh et al., 2015). After that, text data need to be split into training and test data sets and the keyword dictionary is formulated from training data.

2.3 Training & Testing of Machine Learning (NB) Algorithm

A Bayesian method has been used here to construct the classifier. A Naïve Bayes (NB) classifier is used to find the joint probabilities of words and classes within a set of free text. The probability of a class A for a given text field B can be calculated by using Bayes’ law:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

Since $P(B)$ is constant for all classes, only the other variables need to be maximised. It is assumed that classes are independent of each other (Naïve assumption). The classification task is done by considering prior probability information and likelihood of the incoming information to form a posterior probability model of classification. NB model is effectively applied for (Lantz, 2013):

- Text classification such as, email filtering, topic categorization etc.
- Problems in which the information from numerous attributes should be considered simultaneously in order to estimate the probability of an outcome.

The NB classifier is typically trained on data with categorical features. A sparse matrix indicates the frequency of the appearance of each keyword in the bag-of-words for each class in the training data. The training algorithm for the NB classifier can be seen in Algorithm 1. We use a Laplace Estimator (the “1” in Line 8) to nullify the zero-frequency words.

Algorithm 1 Training NB algorithm

$Train\ NB(D_f, D_p)$; D_f = Text field labelled as failure & D_p= Text field labelled as preventive
Extract keywords from $D_f \rightarrow V_f$
Extract keywords from $D_p \rightarrow V_p$
$\boldsymbol{for\ each}\ c \in \{f, p\}\ \boldsymbol{do}$
  $N_c = |D_c|$ No. of documents in class c
  $prior[c] \leftarrow N_c/N$
  $\boldsymbol{for\ each}\ t \in V_c$
  $\boldsymbol{do}\ T_{ct}$ Count occurrences of word t in D_c
  $condprob[t][c] \leftarrow \frac{T_{ct}+1}{\sum_{t'} T_{ct'}+1}$
$\boldsymbol{return}\ prior, condprob$

With the output of Algorithm 1, we can classify new free text fields as failure or preventive maintenance in the following manner. Suppose a free text field contains the words w_1, $w_2, \ldots, w_M$. We may then predict the class label, c^* using (Lantz, 2013)

$$c^* = \arg \max_{c=\{f,p\}} \ prior[c] \prod_{i=1}^{M} condprob[w_i][c]$$

We evaluate the performance of the classifier on the unseen test datasets as is standard practice in machine learning. We employ the following measures to quantify the classifier performance (Prytz, 2015):

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN}$$

Where, TP, TN, FP, FN represent True Positive, True Negative, False Positive and False Negative classifications respectively. Finally the classifier constructed with WON data is applied to DD to classify each downtime event as "failure" or "preventive maintenance".

3 Case Study

Maintenance data coming from coal pulverized mills of an Australian power plant over a 21 year period are used here to illustrate the application of proposed information extraction methodology. The data for 12 mills includes WONs and DD. Fig. 2 shows the process of recording WON and DD during maintenance process and we can see that, the data sets are consistent with our assumptions: DD indicates that mill was actually stopped but does not specify *why*, and the WONs contain more information, but do not indicate if the issue causes a stoppage.

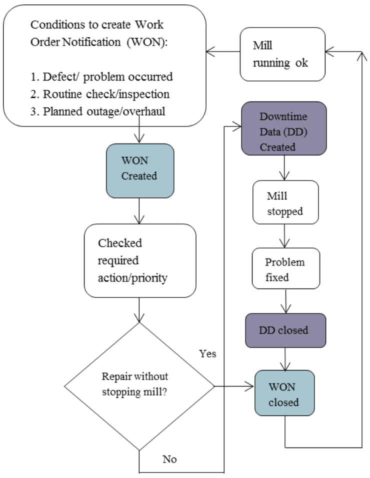

Figure 2 Creation of two data sources during the maintenance process (Coal Mill)

The incompleteness in both of the data sources independently motivates the use of the methodology developed in this paper. After applying the filters (Section 2.1) to WON (total 9401 documents), the frequencies of failure and preventive maintenance are 1068 and 8333. In this analysis, R project is used here to analyse text data as well to train and test the NB model according to the methodology in Section 2. After applying text cleaning process, mentioned in section 2.2, a total of 1582 keywords were identified and were saved to dictionary. The NB classifier is trained on keyword dictionary and the performance of the classifier is tested by comparing predicted values of failure and scheduled maintenance work orders with actual ones not utilized in the training set. Fig. 3 shows the model performance on the test data. Precision (also called positive predictive value) is the fraction of predicted failures that are truly failures, while recall (also known as sensitivity) is the fraction of predicted failures that are identified correctly.

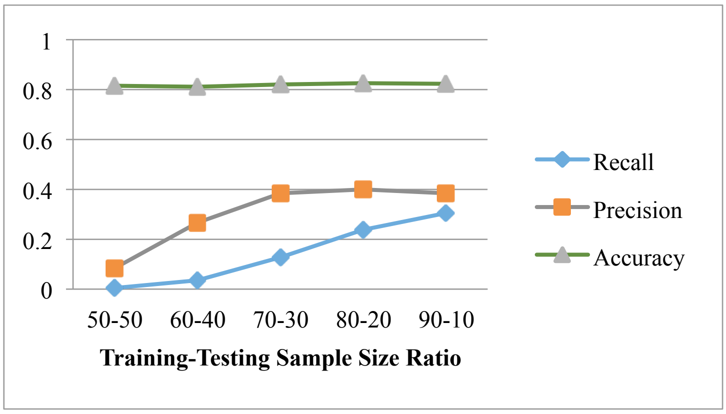

Figure 3 Performance metrics for NB model

The tested NB model is finally applied to DD and labelled them into failure and planned preventive information. Table 1 shows the outcome of the prediction per mill. It is important to mention that the predicted values cannot be validated for this case because there is no evidence of mill failure information for DD. Predicted values can be used to plot cumulative number of failures. For example, Fig. 4 shows cumulative number of failures for mill XA in two different cases. Unplanned WON means all the notifications which are urgent and unplanned while text mining means cumulative number of failures after applying text mining and filters. This information is ready for inclusion into a wide variety of reliability models (Wang and Pham, 2006).

Table 1 Predicted frequencies of failure and preventive maintenance information

	Unit X						**Unit Y**						Row Total
	Mill						Mill						
	A	B	C	D	E	F	A	B	C	D	E	F	
Failure	34	42	48	37	39	37	46	32	53	37	54	33	490
Preventive Maintenance	79	99	95	73	110	69	67	77	101	107	103	82	1064
Column Total	113	141	143	110	149	106	113	109	154	144	157	115	1554

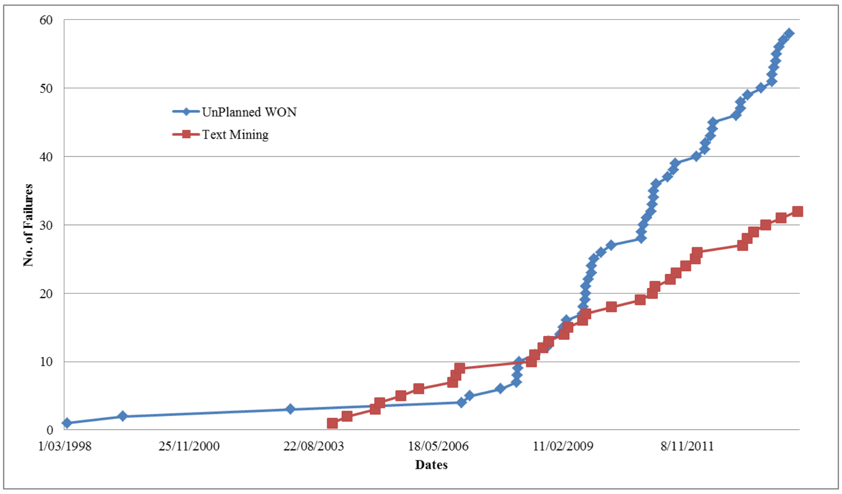

Figure 4 Cumulative number of failures for mil XA over a 16 year period

4 Conclusion

A new data extraction methodology has been proposed here to obtain information for reliability modelling from commonly-recorded asset data. To overcome the incompleteness in maintenance one dataset with reliable free text description was used to construct keyword dictionary and extract failure and preventive maintenance data from another data source with reliable stop time data. To the best of the authors', this is the first use of text mining approaches to extract reliability information from multiple heterogeneous data sources. Such data fusion is a key challenge in exploring Big Data (Wu et al., 2014). The developed classification can be utilized to build reliability models for the optimisation of maintenance and availability of real assets.

References

Alkali, B M., Bedford, T., Quigley, J and Gaw, J. 2009. "Failure and maintenance data extraction from power plant maintenance management databases" *Journal of Statistical Planning and Inference* 139: 1766-1776.

Bastos, P., Lopes, I. and Pires, L. 2014. "Application of data mining in a maintenance system for failure prediction" *Safety, Reliability and Risk Analysis: Beyond the Horizon- Steenbergen et al. (Eds)*: 933-940.

Jeon, J., and Sohn, So Y. 2015. "Product failure pattern analysis from warranty data using association rule and Weibull regression analysis: A case study" *Reliability Engineering & System Safety* 133: 176-183.

Lantz, B. 2013. *Machine learning with R*. Packt Publishing, Birmingham, UK.

Louit, D M., Pascual, R. and Jardine, A K S. 2009. "A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data" *Reliability Engineering & System Safety* 94 (10): 1618-1628.

Noh, H., Jo, Y. and Lee, S. 2015. "Keyword selection and processing strategy for applying text mining to patent analysis" *Expert Systems with Applications* 42: 4348-4360.

Prytz, R., Nowaczyk, S., Rognvaldsson, T. and Byttner, S. 2015. "Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data" *Expert Systems with Applications* 42: 4348-4360.

Wang, W. and Pham, H. 2006. *Reliability & Optimal Maintenance*. Springer, London, UK.

Wu, X., Zhu, X., Wu, G Q. and Ding, W. 2014. "Data mining with big data" *IEEE Transactions on Knowledge and Data Engineering* 15 (8): 796-801.